12.4 二つずつなら独立、全体では独立でない

ワークシート

問題

STEP III 2021、問 12

先に進む前に、まず問題を読んでおこう。問題は、OCR が配布している 2021 年の STEP 過去問一式（zip ファイル）のうち、STEP III の問題用紙にある：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

? **考えてみよう**　$\mathrm{P}(X_{12}=1)$ はいくらか。$\mathrm{P}(X_{12}=1 \text{ かつ } X_{23}=1)$ はいくらか。

段階 1　対ごとの独立（pairwise independence）

扱う内容：小問 (i)、独立性（independence）

? **考えてみよう**　$\mathrm{P}(X_{12}=1)$ はいくらか。また、$\mathrm{P}(X_{12}=1 \text{ かつ } X_{23}=1)$ はいくらか。それらの積と比べてみよ。

段階 2 得点の平均（mean）と分散（variance）

扱う内容：小問 (i)、平均と分散

> **手を止めて、やってみよう** 得点は $\binom{n}{2}$ 個の確率変数の和である。期待値（expectation）はつねに足し合わせることができる。では、和の分散が分散の和に等しくなるのはどのようなときか。対ごとの独立で十分だろうか。

段階 3 分散の公式

扱う内容：小問 (ii)

> **考えてみよう** Y_i の平均がすべて 0 ならば、和の分散はその 2 乗の E に等しい。2 乗を展開せよ。各積 Y_iY_j は何回現れるか。

段階 4　奇数・偶数のゲーム

扱う内容：小問 (iii)、独立でないこと

考えてみよう　Z_{12} の分布（distribution）とその平均を求めよ。$Z_{12}Z_{23}$ が 0 でないのはどのようなときで、そのときの値はいくらか。$\mathrm{E}(Z_{12}Z_{23})$ と $\mathrm{E}(Z_{12})\mathrm{E}(Z_{23})$ を比べよ。

段階 5 **分散**

扱う内容：小問 (iii)、分散

> **手を止めて、やってみよう** 小問 (ii) を用いよ。交差項（cross terms）に寄与するのは、ペアどうしの組のうちどれか。共通のプレイヤーをもたないものか、それともちょうど 1 人を共有するものか。それぞれ何組あるか。

背後にある数学

複数の事象の独立性

事象 $A_1, \ldots, A_n$ が**独立**であるとは、そのどの部分集まりについても、共通部分の確率が各確率の積に等しいことをいう。3 つの事象の場合、これは 4 つの条件を意味する。対についての 3 つの条件と、$\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$ である。セルゲイ・ベルンシュタイン（Sergei Bernstein）は 1920 年代に図 12.4.1（オンライン版の節全体を参照）の例を示し、対についての条件から 3 つについての条件は導かれないことを示した（逆に、3 つについての条件だけからも対についての条件は導かれない）。問題における一致の事象 X_{12}、X_{23}、X_{13} は、ベルンシュタインの硬貨とまったく同じ構造をしている。

無相関は独立ではない

共分散が 0、すなわち $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ であること（無相関（uncorrelated））は、独立性よりさらに弱い。STEP III 2000 Q14（演習問題にある）は、2 つの値しかとらない変数では 2 つの概念が一致するが、3 つの値をとる場合には一致しないことを示している。簡単な例を挙げよう。X を $\{-1, 0, 1\}$ 上の一様分布（uniform distribution）に従う変数とし、$Y = X^2$ とする。すると $\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\mathrm{E}(Y)$ であるが、Y は X によって完全に決まる。共分散が測るのは線形な関連だけなのである。

安価な乱数

乱数を生成することは、計算機にとって高くつく作業である。1979 年、Larry Carter と Mark Wegman は、データを擬似乱数的な位置に格納する**ハッシュ法**（hashing）をはじめとする多くのアルゴリズムでは、対ごとに独立なランダムな選択でも完全に独立なものと同じくらいうまくいくことを示した。対ごとに独立な値は非常に安価に生成できる。たとえば、a と b をランダムに選んで $k = 0, 1, \ldots, p-1$ に対して $a + bk \pmod{p}$ をつくると、どの 2 つも独立な p 個の値が得られる。これがうまくいく理由は、この問題の分散の計算にある。チェビシェフ（Chebyshev）の不等式に必要なのは分散だけであり、分散に必要なのは対だけなのである。

次の一歩

1 **2 つの値。** STEP III 2000、問 14（step.maths.org/questions/00-s3-q14）に挑戦せよ。この問題は、それぞれ 2 つの値しかとらない確率変数では $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ から独立性が従うことを示し、3 つの値をとる場合の反例（counterexample）を求めるものである。

2 **より大きなチーム。** 小問 (i) で $n = 7$ のとき、得点の平均と標準偏差（standard deviation）を求め、例として挙げられた得点 7 と比較せよ。

3 **再び誕生日。** 指示変数を用いて、k 人の中で誕生日が一致する対の個数の平均と分散を求めよ（Section 9.9）。それらの指示変数は対ごとに独立か。

4 **3 つについては成り立つが対ごとには成り立たない。** 確率がいずれも $\frac{1}{2}$ で、$\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$ を満たすが、対ごとに独立ではない 3 つの事象を見つけよ。

さらに学ぶために

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press）。独立性を扱った最初のいくつかの章。
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing***（Cambridge University Press）。対ごとの独立性とそのアルゴリズムへの応用。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press、オンラインで無料公開）。指示変数と共分散について。

詳しい解説付きの解答は第 12.4 節にある：

https://beyondstep.keithmcnulty.org/ja/ch12/pairwise-independence.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com